ISSN 1561-5359. Штучний інтелект, 2018, № 4

УДК 004.89

А.З. Муслимова, Г.Б. Галымжан, С.А. Кудубаева
Костанайский государственный университет имени А. Байтурсынова, Казахстан
ул. Байтурсынова, 47, г. Костанай, 110000

АВТОМАТИЗАЦИЯ СУРДОПЕРЕВОДА ЧЕРЕЗ СИСТЕМУ РАСПОЗНАВАНИЯ ЖЕСТОВ ГЛУХИХ

A.Z. Muslimova, G.B. Galymzhan, S.A. Kudubayeva
Kostanay State University After A. Baitursynov, Kazakhstan
47, Baitursynov St., Kostanay, 110000

AUTOMATION OF SIGN LANGUAGE THROUGH THE SYSTEM OF RECOGNITION OF THE GESTURE OF THE DEAF

В статье рассматриваются методы распознавания жестов глухих, основанные на анализе внешних признаков жеста и анализе трехмерной модели руки. Рассмотрены подходы к обнаружению и распознаванию жестов при реализации технологии интеллектуального человеко-машинного интерфейса. Предложен новый алгоритм на основе детектора Виолы-Джонса, его модификации и метода выделения связанных областей, использован метод корреляционного анализа для распознавания жестов на видеопоследовательности Представлены результаты тестирования работы предложенного алгоритма. Показано, что использование указанного алгоритма дает возможность эффективного распознавания жеста из одного слова на видеопоследовательностях.

Ключевые слова: распознавание жестов, цифровая обработка изображений, компьютерное зрение

This article examines the methods of recognizing gestures of the deaf people, based on the analysis of external signs of gesture and analysis of the three-dimensional model of a hand. The gesture recognition algorithm proposed by us consists of four stages: system training, setting a specific set of gestures and corresponding commands; transformation of the original image into the initial representation (mathematical transformations, the calculation of the principal components); localization and segmentation of hand areas in the image based on key characteristics; classification mechanism (simulation).

Keywords: gesture recognition, digital image processing, computer vision, computer linguistics

Введение

Жестовый язык глухонемых является довольно сложным языком. В нем, как и в естественном языке используются грамматика и правила жестикуляции

Разработка системы автоматического распознавания жестов открывает путь для создания более естественных человеко-машинных интерфейсов, убирает ограничения общения, с которыми сталкиваются глухие люди в повседневной жизни. Исследования по созданию систем автоматического сурдоперевода, в основном, посвящены жестам азбуки ASL. Настоящая работа рассматривает особенности преобразования в текстовый вид жестов глухих казахов. В дальнейшем планируется расширить исследования за счет жестов, которые представляют собой движения рук, головы, губ и обозначают не букву, цифру, слово, а целое выражение или ситуацию.

Основная часть

Современные методы распознавания жестов используют в основном технику на основе компьютерного зрения. Распознавание жестов может осуществляться методом создания пространства кривых, суть которого заключается в нахождении граничных контуров руки глухого, говорящего жестами [1]. Данный подход достаточно надежен и инвариантен к перемещениям и вращениям руки, однако требует больших вычислительных затрат. Известен алгоритм распознавания положения руки с помощью изображений скелета руки [2]. В данном случае применяется многокамерная система для нахождения центра гравитации руки и наиболее отдаленных от него точек, обеспечивая, таким образом, информацию о положении кончиков пальцев, которая используется для построения изображений скелета руки и, соответственно, распознавания. Остальные методы рас-

познавания жестов используют специальные методы сравнения [3], дескрипторы Фурье, нейросети, гистограммы положения, фильтрацию точек [4]. При распознавании используется фиксированный набор жестов, записанных на видео камеру.

Поэтому быстродействие и простота алгоритма имеют большое значение. Такой подход включает сегментацию изображения руки на основе цветовых характеристик кожи и ограничений размерности. Ограничение размерности является необходимым условием, так как большая размерность вызывает огромные вычислительные затраты [5]. Все движения руки описываются с помощью базисных векторов. Для уменьшения размерности пространства наблюдаемых векторов без существенной потери информации рекомендуется применять анализ главных компонент (АГК), а для отображения характерных особенностей – анализ независимых компонент (АНК) [6]. Для того чтобы система отвечала нужными реакциями на определенные совокупности внешних воздействий, необходимо подключить процесс обучения, который заключается в адаптации системы к конкретным движениям рук пользователя (заданному набору жестов). В качестве объектов обучения выступают визуальные изображения рук. Итак, на основе этих предварительных процедур обработки генерируется сигнал, который несет информацию о жесте на изображении. Далее жест сравнивается с набором жестов из базы данных и, в случае успешной классификации, ему присваивается определенная команда. На выходе системы формируется управляющий сигнал, предающий команду, на основе которой автоматическое устройство выполняет то или иное действие.

Предлагаемый нами алгоритм распознавания жестов состоит из следующих этапов: 1) обучение системы, задание конкретного набора жестов и соответствующих команд; 2) преобразование исходного изображения в начальное представление (математические преобразования, вычисление главных компонент); 3) локализация и сегментация областей руки на изображении на основе ключевых характеристик; 4) механизм классификации (моделирования).

Для обработки изображения проводим сегментацию изображения руки на основе ключевых характеристик. В качестве признака, который используется для отделения руки от фона на изображении, можно использовать цвет кожи. В данном случае для реализации сегментации применяется пиксельная модель кожи. Модель формируется исходя из информации о цветности (тон и насыщенность), полученная в результате предварительного обучения, которое непосредственно состоит в размещении руки пользователя в области так называемого обучающего квадрата. Пиксели, заключенные в эту область, используются для обучения модели, после чего выделенные пиксели преобразовываются из цветового пространства RGB в пространство HSL, откуда затем получается информация о цветности. Значения цветового тона H и насыщенности S для каждого выбранного пикселя образуют набор $\underset{x}{\rightarrow}\left(\underset{x_1}{\rightarrow},\underset{x_2}{\rightarrow}\ldots\underset{x_n}{\rightarrow}\right)$, где n – количество отсчетов (пикселей), $\underset{x_i}{\rightarrow}=(H_i,S_i)$ – значения цветового тона и насыщенности i-го пикселя. Для представления функции плотности вероятности, описывающей принадлежность пикселей к цвету кожи, выбрана гауссова функция плотности вероятности (ГФПВ).

В конечном итоге результатом процесса сегментации будет представление руки в виде реперной области путем применения алгоритма объединения пикселей. Полученные таким образом результаты являются инвариантными к фону и различным условиям освещения.

Алгоритм можно модифицировать, если включить дополнительные этапы предобработки, такие как калибровка камер, фильтрация и т.п. Этап сегментации

достаточно прост и при использовании в сложных условиях работы нуждается в усовершенствовании. Надежное выполнение алгоритма распознавания жестов предполагает учет неоднозначной природы статических и динамических жестов, проблем выделения руки на изображении, условий освещения и помех.

Согласно формализации Stokoe каждый жест языка глухонемых можно представить в виде совокупности пяти компонент:

- TAB – позиция руки или рук,
- DEZ – конфигурация руки или рук,
- ORI – ориентация руки,
- SIG – движение руки, пальца и т.д.,
- выражение лица-тела.

На основе указанной формализации была предложена письменная форма американского жестового языка (ASL) и в последующем разработаны системы обозначений Hamburg Notation System и SignWriting, которые использовались для документации ASL. В настоящей работе будем опираться только на концепции, изложенные в этих системах. В системе SignWriting имеется 261 конфигураций руки. Под конфигурацией (формой) руки здесь понимается установленное положение пальцев и кисти руки.

Задача автоматического перевода жестового языка включает слежение за движением руки, распознавание последовательности показанных жестов и их отображение в текстовую форму. Если обозначить через T^* - сегменты видеоряда, в каждом из которых S показывается отдельный жест, то задачу распознавания жестов можно представить в виде поиска функции $R: T^* \to S \times M$, которая каждому сегменту из T^* сопоставляет элемент из $S \times M$. Несмотря на простую формулировку, задача является довольно сложной по некоторым причинам:

- жесты, которые трудно отличить даже человеку, некоторые жесты схожи;
- скорость жестикуляции может отличаться при разных показах жестов;
- размер ладони у разных людей существенно отличается;
- следует учитывать такие факторы, как разные цвета кожи у людей и изменения освещенности;
- распознавание жестов необходимо осуществить в реальном времени;
- несмотря на правила, каждому человеку свойственны свои особенности жестов;
- как правило, жестикуляция производится плавно и слитно, что затрудняет процесс сегментации показанных жестов.

Система перевода жестового языка глухонемых в текстовый вид, получая на входе кадры видеоряда, вычисляет ключевые характеристики жеста в каждом кадре, сегментирует видеоряд на сегменты, каждый из которых соответствует отдельно показанному жесту и после распознается сам жест. На сегодняшний день в свободном доступе имеются библиотеки и программные платформы, которые могут быть использованы для решения части исследуемых задач.

Одним из перспективных методов распознавания образов считается метод Виолы-Джонса – алгоритм, позволяющий обнаруживать объекты на изображениях в реальном времени. Этот метод предложен в 2001 году Полом Виолой и Майклом Джонсом. Данный метод является основополагающим для поиска объектов на изображении в реальном времени в большинстве существующих алгоритмов распознавания и идентификации. Так же он является одним из лучших по соотношению эффективности распознавания и скорости работы. Алгоритм показывает отличные результаты и распознает объекты под небольшим углом, примерно до 30 градусов, и при различных условиях освещенности. Метод Виолы-Джонса является одним из лучших по соотношению показателей эффективность распознавания/скорость работы. На момент написания алгоритм является основополагающим для поиска объектов на изображении в реальном времени в большинстве

существующих методов распознавания и идентификации.

Основные понятия метода Виолы-Джонса:

1. Представление изображения в интегральном виде, что позволяет вычислять быстро необходимые объекты.
2. Использование признаков Хаара, с помощью которых происходит поиск нужного объекта (в данном контексте, лица и его черт).
3. Бустинг (от англ. boost- улучшение, усиление) для выбора наиболее подходящих признаков для искомого объекта на данной части изображения.
4. Использование каскадов признаков для быстрого отбрасывания окон, где не найдено лицо.

В результате анализа существующих решений можно сделать вывод, что специализированного классификатора, предназначенного для распознавания поз руки и удовлетворяющего указанным требованиям, еще не существует. Самым известным классификатором позы считаются детектор Виолы-Джонса и его модификации, которые применяются во многих работах. Недостатки этого детектора заключаются в длительном времени обучения, требовании к большой обучающей базе изображений, и чувствительности к наклонам и деформированию объекта [5], что делает данный метод не самым подходящим решением для задачи распознавания позы руки в данной работе. Однако данный детектор является наилучшим кандидатом для задачи обнаружения руки на видеокадрах в режиме реального времени, благодаря его точности и скорости работы [6].

Таким образом выбора оптимальных алгоритмов и методов обнаружения и распознавания жестов руки на видео в режиме реального времени разработан усовершенствованный алгоритм, который позволяет уменьшить вероятность ложных обнаружений и повысить быстродействие работы системы.

Данный алгоритм основан на методе Виолы-Джонса с использованием метода выделения связанных областей и корреляционного метода состоит из следующих этапов:

1. Загрузка видео файла.
2. Анализ кадров.
3. Изменение размера изображения.
4. Перевод изображения в полу-тоновое.
5. Бинаризация изображения.
6. Пороговое преобразование.
7. Операции морфологической обработки изображения.
8. Выделение связанных областей.
9. Использование метода Виолы-Джонса.
10. Поиск руки на изображении.
11. Получение матрицы эталонов.
12. Проход по всем эталонам и вычисление коэффициента корреляции. При совпадении эталона с изображением достигает максимального значения, равного единице.
13. Распознавание жеста

Экспериментальная часть

Алгоритм на основе детектора Виолы-Джонса, его модификации и методов выделение связанных областей и корреляционный, предназначенный для распознавания двуручных жестов на видеопоследовательности. Эксперимент был проведен с помощью специализированного пакета для решения инженерных, научно-технических и экономических задач Math Works MATLAB R2018a.

Читаем видео файл и запускаем детектор лица. Сначала зададим детектор каскадного объекта vision.CascadeObjectDetector, который определяет местоположение лица и\или его частей (носа, глаз или рта) в видеокадре.

Детектор каскадного объекта использует алгоритм обнаружения Виолы-Джонса и подготовленную классификационную модель для обнаружения. По умолчанию детектор настроен на обнаружение лиц, но его можно использовать для обнаружения других типов объектов.

References

1. Davis, J., Shah, M. Visual Gesture Recognition (1994). EE. Proc.-Vis. Image Signal Process. – April 1994. –Vol. 141, № 2.

2. Chang, C.-C., Chen, I.-Y., Huang, Y.S. Hand Pose Recognition Using Curvature Scale Space (2002). IEEE International Conference on Pattern Recognition.
3. Utsumi, A., Miyasato, T., Miyasato, F. Miyasato Multi-Camera Hand Pose Recognition System Using Skeleton Image (1995). IEEE International Workshop on Robot and Human Communication. P. 219-224.
4. Rosales, R., Athitsos, V., Sigal, L., Sclaroff, S. 3D Hand Pose Reconstruction Using Specialized Mappings (2001). IEEE International Con. on Computer Vision. P. 378-385.
5. Manresa, C. Hand tracking and gesture recognition for human-computer interaction (2005). Electronic Letters on Computer Vision and Image Analysis. № 5(3). P. 96-104.

RESUME

A.Z. Muslimova, G.B. Galymzhan, S.A. Kudubayeva

Automation of sign language through the system of recognition of the gesture of the deaf

Intellectual technologies for sign languages (including Kazakh) and their features have not been considered properly in world science yet, therefore the development of such technologies is more relevant than ever. Also, the Kazakh sign language is subject to partial extinction, and this requires special attention from linguists, sign language interpreters and software developers.

For the selection of optimal algorithms and methods for detecting and recognizing hand gestures in real-time video has been developed an improved algorithm that reduces the likelihood of false detections and improves the speed of the system.

The undoubted advantage of the approach is its universality in terms of taking into account the diversity of possible conditions arising while recognition of visual objects.

The application of the Viola-Jones algorithm with the use of correlation methods for image recognition yielded good results – the recognition of most gestures varies in the range of 80-96%. In our work for image processing, we carry out segmentation of the hand image on the basis of key characteristics. As a sign, which is used to separate the hands from the background in the image, the skin color of the person on the video stream is used. Moreover, to implement the segmentation the pixel skin model is applied. The model is formed on the basis of information about chromaticity (tone and saturation), obtained as a result of preliminary training, which directly consists in placing the user's hand in the area of the so-called learning square. The pixels enclosed in this area are used to train the model, after which the selected pixels are converted from RGB color space to HSL space, from which chroma information is then obtained. A Gaussian probability density function (GPDF) is chosen to represent the probability density function describing the pixels belonging to skin color.

The subject of research is the task of creating automatic system of a sign language interpretation and systems equipped with a more natural human-machine interface for deaf people. We managed to achieve high processsing speed and simplicity in implementation, but algorithms and methods based on skin color have major drawbacks: they cannot distinguish objects with similar color (for example, face, hands, and other parts of the body); their efficiency strongly depends on the illumination and decreases sharply when it changes. This article discusses methods for recognizing gestures of deaf people based on the analysis of external signs of a gesture and the analysis of a three-dimensional model of a hand. The main task is to develop algorithms for recognizing one-handed and two-handed gestures of the deaf and developing educational software applications for learning sign language. Gesture recognition methods that use special comparison methods such as Fourier descriptors, neural networks, position histograms, point filtering are considered. Fixed set of gestures recorded on a video camera, and this set is taken as the standard. The Mathlab program recognizes gestures with fairly good accuracy. The gesture recognition algorithm proposed by us consists of four stages: system training, setting a specific set of gestures and corresponding commands; transformation of the original image into the initial representation (mathematical transformations, the calculation of the principal components); localization and segmentation of hand areas in the image based on key characteristics; classification mechanism (simulation).

Надійшла до редакції 10.10.2018